Brazilian Journal of Biology®

ISSN 1519-6984 (Print)
ISSN 1678-4375 (Online)

Notes and Comments

ImportanceIndice: an R package for application of the percentage of importance indice

G. L. Demolin-Leite[a]* and A. M. Azevedo[a]

[a]Universidade Federal de Minas Gerais – UFMG, Instituto de Ciências Agrárias – ICA, Montes Claros, MG, Brasil

Events can have different magnitudes, frequencies, and distributions of occurrence. The problem can be worse or the solution better if greater frequencies and magnitudes are presented with aggregated distribution in the production system (Demolin-Leite, 2021, 2024). The Percentage of Importance Indice (*% I.I.*) based on this triplet to identify loss and solution sources, classifying them according to their importance in terms of loss or income gain, on the productive system (Demolin-Leite, 2021, 2024). The *% I.I.* can be significant for preserving native areas, avoiding their degradation, assisting the traditional communities, like the *quilombolas* (rebellious slaves refuge area in the Brazilian colonial period), indigenous, collectors (e.g., fruits), to identify the true loss sources of production in native plants. Thus, with the help of extension researchers, they can plan the best management of these potential pests, making more money, avoiding tree-cut for charcoal production (Demolin-Leite, 2024). This index and its derivations (e.g., *non attention level*) were obtained using the statistical programs Biodiversity Professional program, version 2 (Krebs, 1989) – for chi-square test - and System for Analysis Statistics and Genetics, version 9.1 (UFV, 2007) – for simple regression analysis -, and also part of the calculations (e.g., maximum estimated production) using an Excel datasheet. However, the transfer of information from the data obtained via the statistical programs mentioned above, as well as the calculations performed using the Excel datasheet, in addition to being labor intensive, could incur mathematical errors due to the volume of equations and data. For this purpose, a package and its manual were developed, via the R program, to perform the statistics and calculations necessary to obtain the *%I.I.* and its derivations (Demolin-Leite and Azevedo, 2022). This study aimed to demonstrate to use of the R-Package “Importance Indice” (Demolin-Leite and Azevedo, 2022) using adapted published data (simplified) (see Demolin-Leite, 2024) in relation to those obtained with the statistical programs mentioned above. The package is available on Cran’s platform (Demolin-Leite and Azevedo, 2022).

The equation of the % of Importance Indice (*% I.I.*) (Demolin-Leite, 2021, 2024) is: $\% I.I. = \{(ks_1 \times c_1 \times ds_1)/\Sigma (ks_1 \times c_1 \times ds_1) + (ks_2 \times c_2 \times ds_2) + (ks_n \times c_n \times ds_n)\}\times 100$, where:

i) *key source* (*ks*) is: *ks* = reduction on production (*R.P.*)/ total *n* of the *L.S.* or effectiveness of the solution (*E.S.*)/ total *n* of the *S.S.*. Where *R.P.* or *E.S.* = $R^2 \times (1 - P)$ when it is of the first degree, or *R.P.* or *E.S.* = $((R^2 \times (1 - P)) \times(\beta_2/\beta_1)$ when it is of the second degree. Where, R^2 = determination coefficient and *P* = significance of ANOVA, β_1 = regression coefficient, and β_2 = regression coefficient (variable²), of the simple regression equation of the *L.S.* or *S.S.* (Table 1). When a *S.S.* acts on more than one *L.S.*, theirs *E.S.* are summed. *E.S.* or *R.P.* = 0 when *E.S.* or *R.P.* is non-significant on the *L.S.* or *R.P.*, respectively. Simple equations were selected by observing the criteria: 1) distribution of data in the figures (linear or quadratic response), 2) the parameters used in these regressions were the most significant ones ($P \leq 0.05$), 3) $P \leq 0.05$ and *F* of the Analysis of Variance of these regressions, and 4) the coefficient of determination of these equations (R^2).

ii) *constancy* (*c*) is (Demolin-Leite, 2021): *c* = Σ of occurrence of *L.S.* or *S.S.* on the samples. Where, absence = 0 or presence = 1 (Table 1).

And

iii) *distribution source* (*ds*) (Demolin-Leite, 2021) is: *ds* = 1 - *P* of the chi-square test of *L.S.* or S.S. on the samples. The type of distribution (aggregated, random, or regular) of *L.S* or *S.S.* was defined by the Chi-square test (Table 1).

These data, above, are obtained, by R-package (Demolin-Leite and Azevedo, 2022) (Chart 1).

Percentage of loss of production per loss source (*% L.P.L.S.*) is: *% L.P.L.S.* = (*L.P.L.S./M.E.P.*) x 100, where:

a) Loss of production per loss source (*L.P.L.S.*) = total *n* of the *L.S.* x *R.P.* of the *L.S.*, and maximum estimated production (*M.E.P.*) = Total production (*P.*) + $\Sigma\ L.P.L.S._1 + ...L.P.L.S._n$.
b) Income gain (*I.G.*) = *L.P.L.S.* x *E.S.* and *% I.G.* = (*I.G.*/ *M.E.P.*)*100. In this case, the *E.S.* of the *S.S.* is separated per *L.S.* (Demolin-Leite, 2024).
c) Attention level (*A.L.*) = (*n* of the *L.S.* per sample x 0.75)/*% L.P.L.S.* and 0.75 = 1% of loss fruits x 0.75 (safety margin), and
d) Non-attention level (*N.A.L.*) = (*A.L.* x 1.25)/*E.S.* and 1.25 = 25% plus as safety margin (Demolin-Leite, 2024).

These data, above, are obtained, by R-package (Demolin-Leite and Azevedo, 2022) (Chart 2).

*e-mail: germano.demolin@gmail.com
Received: April 28, 2022 – Accepted: September 07, 2022

Table 1. Aggregated (Agg.), regular (Reg.), or random (Ran.) distribution (Dist.) of the loss (*L.S.*) or solution sources (*S.S.*); and simple regression equations with their coefficients of determination (**R²**), significance (***P***) and ***F*** of the analysis of variance (Var.) (ANOVA) of reductions of total fruit production (F.P.) by *L.S.* and reductions of these *L.S.* by *S.S.* on 20 samples.

Loss sources	**Chi-square test**			
	Var.	**Mean**	*P*	**Dist**
$L.S._1$	123.00	10.50	0.00	Agg.
$L.S._2$	5.04	1.90	0.00	Agg.
$L.S._3$	23.01	4.20	0.00	Agg.
$L.S._4$	1099.73	37.05	0.00	Agg.
Solution sources				
$S.S._1$	0.64	0.70	0.56	Ran.
$S.S._2$	359.00	24.45	0.00	Agg.
$S.S._3$	2.03	1.65	0.22	Ran.
Simple regression analysis		**ANOVA**		
		R^2	*F*	*P*
$F.P.=23.54+193.38xL.S._1-4.10xL.S._1^2$		0.90	75.78	0.0001
$F.P.=1927.97-423.27xL.S._2$		0.72	47.04	0.0001
$F.P.=141.56+397.26xL.S._3-17.38xL.S._3^2$		0.71	21.06	0.0001
*$F.P.=330.55+188.86x\ L.S._3$		0.66	34.49	0.0001
$L.S._1=3.40+25.15xS.S._1-9.55xS.S._1^2$		0.44	6.79	0.0068
$L.S._1=-8.17+1.31xS.S._2-0.014xS.S._2^2$		0.80	34.52	0.0001
$L.S._2=2.91-1.44xS.S._1$		0.27	6.49	0.0202
$L.S._2=4.12-0.09xS.S._2$		0.59	25.74	0.0001
$L.S._3=1.70+12.91xS.S._1-5.94xS.S._1^2$		0.42	6.12	0.0100
$L.S._3=-3.65+0.65xS.S._2-0.0085xS.S._2^2$		0.60	12.66	0.0004

*In this case, despite of positive sign, we can use due to this *L.S.* attacks directly and to damages completely the product.

Chart 1. Steps used to obtain the above data.

data("DataLossSource")
Description: example with data from loss sources with four loss sources, one in each column.
data("DataSolutionSource")
Description: example with data from solution sources with three solution sources, one in each column.
data("DataProduction")
Description: example with production data.
data("DataNumberSamples")
Description: example with data of number samples. A data frame with the number of evaluations performed on each individual, the number of months evaluated and the number of evaluations performed per month.
Distribution_LossSource(DataLossSource)
Description: Indicates the distribution of loss sources: aggregate, random, or regular.
Distribution_SolutionSource(DataSolutionSource)
Description: Indicates the distribution of solution sources: aggregate, random, or regular.

The loss sources (*L.S.*, e.g., insect pests), per individual, $L.S._2$, $L.S._3$, and $L.S._1$, showed the highest *% I.I.* (63.23, 36.38, and 0.39%, respectively) on samples. The total number of products loss (e.g., fruits), percentage of production reduction, and attention levels, respectively, per *L.S.*, on 20 samples were: $L.S._3$ = 55.19, 0.25%, and 0.12/sample; $L.S._2$ =

27.48, 0.12%, and 0.11/sample; and *L.S.*$_1$ = 4.00, 0.02%, and 4.11/sample; totalizing on 86.68 lost products and 0.38% of production reduction (Tables 2 and 3).

The effective solution sources (*S.S.*, e.g., natural enemies), per individual, *S.S.*$_1$ and *S.S.*$_2$ showed the highest *% I.I.* (88.68 and 11.32%, respectively) on 20 samples. The *S.S.*$_1$ reduced

Chart 2. Steps used to obtain the above data.

LS=LossSource(DataLoss = DataLossSource,DataProd = DataProduction,verbose = TRUE)
LS
LP=LossProduction(Data=DataLossSource,Prod = DataProduction, Evaluation=DataNumberSamples,SegurityMargen=0.75,MaximumToleranceOfLossFruits=1)
ES<-EffectivenessOfSolution(DataLossSource=DataLossSource, DataSolutionSource=DataSolutionSource,Production =DataProduction)
Id<-SelectEffectivenessOfSolution(ES) # id<-c(TRUE, TRUE, TRUE, FALSE, TRUE)
SS<-SolutionSource(SolutionData = DataSolutionSource, EffectivenessOfSolution = ES,Production = DataProduction,Id = id)
NAL<-NonAttentionLevel(EffectivenessOfSolution = ES,LossProduction = LP,Id = id,Verbose=TRUE)

Table 2. Total number (*n*), reduction on production (*R.P.*), effectiveness of the solution (*E.S.*), key-source (*ks*), constancy (*c*), distribution source (*ds*), number of importance indice (*n.I.I.*), sum of *n. I.I.* (Σ*n.I.I.*), and percentage of *I.I.* by loss source (*L.S.*) or solution source (*S.S.*) by *L.S.* on 20 samples.

Loss sources – R program function: LossSouce								
L.S.	***n***	***R.P.***	***ks***	***c***	***ds***	***n.I.I.***	**Σ*n.I.I.***	***%I.I.***
L.S.$_1$	210	0.019	0.000091	13	1.00	0.001	0.301	0.39
L.S.$_2$	38	0.723	0.018947	10	1.00	0.189	0.301	63.23
L.S.$_3$	84	0.657	0.007822	14	1.00	0.110	0.301	36.38
L.S.$_4$	741	0.000	0.000000	19	1.00	0.000	0.301	0.00
Solution sources – R program function: SolutionSource								
S.S.	***n***	***E.S.***	***ks***	***c***	***ds***	***n.I.I.***	**Σ*n.I.I.***	***%I.I.***
S.S.$_1$	14	0.618	0.044135	10	0.44	0.194	0.219	88.68
S.S.$_2$	489	0.606	0.001240	20	1.00	0.025	0.219	11.32
S.S.$_3$	33	0.000	0.000000	16	0.78	0.000	0.219	0.00

Table 3. Loss sources (*L.S.*), loss of production by loss source (*L.P.L.S.*), % of *L.P.L.S.*, and attention level (*A.L.*); effectiveness of the solution (*E.S.*) per solution source (*S.S.*), income gain (*I.G.*) and its %, and non-attention level (*N.A.L.*) by *S.S* and partial and total sum (Σ) on 20 samples.

Loss of production by loss source – R program function: LossProduction				
L.S.	***L.P.L.S.***	***% L.P.L.S.***	***n per sample***	***A.L.***
L.S.$_1$	4.000	0.0177	0.0972	4.1129
L.S.$_2$	27.48	0.1218	0.0176	0.1083
L.S.$_3$	55.19	0.2446	0.0389	0.1192
Total Σ	**86.6760**	**0.3842**	---	---
Increase in production per solution source and total – R program function: NonAttentionLevel				
		***L.S.*$_1$**		
S.S.	***E.S.***	***I.G.***	***% I.G.***	***N.A.L.***
S.S.$_1$	0.1674	0.6698	0.0030	30.7038
S.S.$_2$	0.0087	0.0349	0.0002	588.7065
Σa	**---**	**0.7047**	**0.0032**	**---**
		***L.S.*$_2$**		
S.S.$_1$	0.2597	7.1367	0.0316	0.5214
S.S.$_2$	0.5884	16.1710	0.0717	0.2301
Σb	**---**	**23.3077**	**0.1033**	**---**
		***L.S.*$_3$**		
S.S.$_1$	0.1908	10.5291	0.0467	0.7812
S.S.$_2$	0.0079	0.4344	0.0019	18.9343
Σc	**---**	**10.9635**	**0.0486**	**---**
Total Σ=Σa+Σb+Σc	**---**	**34.9760**	**0.1550**	**---**

Maximum Estimated Production (M.E.P.) = Total production (P.) + Σ L.P.L.S.$_1$ *+ ....L.P.L.S.*$_n$*. M.E.P.* = 22,561.68 (22,475 + 86.68).

production loss per *L.S.*$_1$, *L.S.*$_2$, and *L.S.*$_3$ (0.67, 7.14, and 10.53 total saved product – *I.G.*, respectively) increasing in % of income gain (0.003, 0.032, and 0.047%, respectively) on productive system. The *S.S.*$_2$ decreased production loss per *L.S.*$_1$, *L.S.*$_2$, and *L.S.*$_3$ (0.04, 17.17, and 0.32 total saved product – *I.G.*, respectively) increasing in % of income gain (0.0002, 0.0717, and 0.0019%, respectively) on productive system. The total reduction in production loss due to loss sources was 34.98 total saved product, with an increase in system productivity of 0.16% due to the solution sources cited above. The non-attention levels (*N.A.L.*) for these *S.S.* were: i) *S.S.*$_1$= 30.70 for *L.S.*$_1$, 0.52 for *L.S.*$_2$, and 0.78 for *L.S.*$_3$, and i) *S.S.*$_2$= 588.71 for *L.S.*$_1$, 0.23 for *L.S.*$_2$, and 18.93 for *L.S.*$_3$ on one plant part evaluated/tree (Tables 2 and 3).

These numbers, obtained using the R-Package "Importance Indice" (Demolin-Leite and Azevedo, 2022), were the same as those obtained with the adapted data (reduced) of the published paper (Demolin-Leite, 2024). However, the R-Package "Importance Indice" (Demolin-Leite and Azevedo, 2022) was faster, more practical, and safer way than those obtained previously via the statistical programs and Excel datasheet mentioned above.

References

DEMOLIN-LEITE, G.L., 2021 [viewed 2 May 2018]. Importance indice: loss estimates and solution effectiveness on production. *Canadian Journal of Agricultural Science* [online], vol. 55, no. 2, pp. 95-101. Available from: http://scielo.sld.cu/pdf/cjas/v55n2/2079-3480-cjas-55-02-e10.pdf

DEMOLIN-LEITE, G.L., 2024. Do arthropods and diseases affect the production of fruits on *Caryocar brasiliense* Camb. (Malpighiales: caryocaraceae)? *Brazilian Journal of Biology = Revista Brasileira de Biologia*, vol. 84, pp. e253215. http://dx.doi.org/10.1590/1519-6984.253215.

DEMOLIN-LEITE, G.L. and AZEVEDO, A.M., 2022. *Package "ImportanceIndice": analyzing data through of percentage of importance indice and its derivations. Manual package* [online]. Vienna: R Foundation for Statistical Computing, 17 p. Available from: https://cran.r-project.org/package=ImportanceIndice

KREBS, C.J., 1989 [viewed 2 May 2018]. *Bray-Curtis cluster analysis* [online]. Available from: http://biodiversity-pro.software.informer.com/

UNIVERSIDADE FEDERAL DE VIÇOSA – UFV, 2007 [viewed 30 June 2018]. *Sistema para Análises Estatísticas e Genéticas (SAEG). Version 9.1* [online]. Viçosa: UFV. Available from: http://arquivo.ufv.br/saeg/